DOCUMENT RESUME

ED 294 892 TM 011 522

AUTHOR Ferrara, Steven; And Others
TITLE Local Assessment Responses to a State-Mandated Minimum-Competency Testing Program: Benefits and Drawbacks.
PUB DATE 31 Mar 88
NOTE 31p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 6-8, 1988).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Citizenship Education; Computer Assisted Testing; Educational Assessment; *Graduation Requirements; High Schools; Mathematics Tests; Microcomputers; *Minimum Competency Testing; Reading Tests; School Districts; *State Programs; *Student Attitudes; *Teacher Attitudes; Writing Evaluation
IDENTIFIERS *Maryland Functional Testing Program

ABSTRACT

A study was undertaken to describe assessment activities of four school districts in Maryland (Washington, Cecil, Montgomery, and Charles Counties) designed to parallel a state-mandated competency-testing program required for high school graduation and to report uses of scores and positive and negative impacts from assessment activities. The context of the study was the Maryland Functional Testing Program, which covers reading, writing, citizenship skills, and mathematics. The Montgomery County school system also administered a computerized adaptive functional mathematics assessment test via microcomputer. Data were collected via telephone surveys of testing directors and impact analysis of students who took the tests. Results indicate that: (1) teachers were enthusiastic about participating in item development and essay scoring; (2) increased emphasis on basic skills objectives, to the detriment of other learning outcomes, was exacerbated by parallel assessments; and (3) further consideration should be given to state involvement in the development and administration of competency tests. Five tables are included. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED294892

DRAFT

LOCAL ASSESSMENT RESPONSES TO A STATE-MANDATED MINIMUM-COMPETENCY TESTING PROGRAM: BENEFITS AND DRAWBACKS

Steven Ferrara - Maryland State Department of Education
Carolyn Seburn - Washington County (Md) Public Schools
Frank Slaughter - Cecil County (Md) Public Schools
Jose Stevenson - Montgomery County (Md) Public Schools
Joseph Willhoft - Charles County (Md) Public Schools

March 31, 1988

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

STEVEN FERRARA

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper to be presented at the annual meeting of the National Council on Measurement in Education, April 6, 1988, New Orleans, LA

TM 011 522

BEST COPY AVAILABLE

LOCAL ASSESSMENT RESPONSES TO A STATE-MANDATED MINIMUM-COMPETENCY TESTING PROGRAM: BENEFITS AND DRAWBACKS

Steven Ferrara - Maryland State Department of Education
Carolyn Seburn - Washington County (Md) Public Schools
Frank Slaughter - Cecil County (Md) Public Schools
Jose Stevenson - Montgomery County (Md) Public Schools
Joseph Willhoft - Charles County (Md) Public Schools

March 31, 1988

Purpose and Rationale

The purposes of this study are to (1) Describe assessment activities of four school districts designed to parallel a state-mandated competency-testing program required for high school graduation; and (2) Report uses of scores and positive and negative impacts from these assessment activities.

Recent analyses of the impacts of state-mandated minimum-competency testing programs have focused only on the effects of such programs on curriculum and instruction (for example, Anderson & Hough, 1986; Corbett & Wilson, 1987, March) and on student achievement (for example, Corbett & Wilson, 1987, March; Popham, 1987). Analyses from the early 1980s, when minimum-competency testing programs were first being planned or implemented, focused on similar concerns. Little attention has been paid to assessment-type responses to state-mandated high-stakes tests. Three exceptions include a discussion of the possibility that local assessments that parallel the National Assessment of Educational Progress (NAEP) could arise in response to the proposed use of NAEP for interstate comparisons (Ferrara & Thornton, in press); the inclusion of two items that focus on practice testing and revisions to local testing programs in a 78-item questionnaire administered in local school districts in two states with mandated minimum-competency testing programs (Corbett & Wilson, 1987, March); and a teacher survey regarding current levels of emphasis in schools on standardized achievement testing (R. Stake, personal communication, December 14, 1987).

In this paper we describe tests designed and implemented by local school district staff in four districts in Maryland that parallel state tests by following abbreviated test specifications and state test administration and scoring procedures. These local tests are referred to here as local "parallel assessments." We also report the results of a telephone survey of testing directors in the 20 other local districts in Maryland regarding (1) Currently implemented or planned local

parallel assessment activities; and (2) Actual or expected score uses, benefits, and drawbacks from local parallel assessments. Finally, we briefly describe an attempt to identify impacts of parallel assessment activities on districts' grade 9 state test pass-rates.

For the purposes of this study we define a local test as parallel to its state testing program counterpart if:

1. Test content (items and content domains) is produced directly from abbreviated test specifications available in documents produced by the State Department of Education;

2. All phases of testing procedures (i.e., from development to reporting and interpretation) follow state competency testing program guidelines to a significant degree; and

3. The test is implemented district-wide.

Context

The context of this study is the Maryland Functional Testing Program (MFTP) which is comprised of four basic skills competency tests. All four tests are criterion-referenced tests generated using test specifications that are based on the state's basic skills curriculum. Passing scores for each of the four tests were set using popular standard setting procedures that involve judgments by various constituencies in the state and data on pass-rates that would result from various passing scores. The Maryland Functional Reading and Mathematics tests and the Maryland Test of Citizenship Skills are typical paper-pencil, multiple-choice tests. The Maryland Writing Test contains two essays and, beginning in 1989, a multiple-choice test of writing conventions. The essays are scored by a professional scoring contractor using Maryland's four-point, modified holistic scoring rubric and training materials developed by Maryland writing assessment specialists. Students in the graduating class of 1989 are required to pass all four tests in order to receive a Maryland high school diploma.

In order to understand the impacts of the statewide Functional Testing Program on one aspect of school district assessment activities, we set out to answer two questions: (1) How many of the 24 districts have implemented parallel assessments and in which content areas? And (2) What are the intended purposes and goals of implementing these parallel assessments? We have framed the second question in terms of the ways in which parallel assessment scores are

used and benefits and drawbacks of implementing these assessments. Our procedures for answering these questions are described below.

Method

Descriptions of Four Local Parallel Assessment Programs

Testing directors/staff members from four school districts in Maryland describe their district's parallel assessment activities: a direct writing assessment in Washington County; citizenship and writing tests in Charles County; mathematics, citizenship, and writing tests in Cecil County; and computer-adaptive testing of mathematics in Montgomery County. These district testing programs represent some of the longest-running local parallel assessments in the state and are located in districts which are fairly representative of geographical regions, SES levels, and school district sizes in the state. Each description covers development, participation, and scoring and reporting activities and district staff reports of score uses and impacts. The purpose of these somewhat detailed descriptions is to provide the reader with a strong sense of the nature of established local parallel assessments. The writers' experience with local parallel assessments also provided the basis for the score use and impact questions used in subsequent interviews (described below).

Telephone Survey

The five authors of this paper interviewed district testing directors to determine the extent of parallel assessment activities in the state, features of existing parallel assessment programs, and actual or expected uses of scores and impacts (i.e., benefits and drawbacks) of local parallel assessments.

Sample, instrumentation, and procedures. Testing directors in the other 20 districts in Maryland were surveyed during January, 1988 using one of two interview schedules. Both schedules are accompanied by an introductory script in which the existence of local parallel assessments was determined, the purposes of the study and the definition of parallel assessment were communicated to the respondent, and a subsequent interview was scheduled. Also, both schedules contain questions regarding actual or hypothetically expected uses of scores and impacts (i.e., possible benefits and drawbacks) from parallel assessments.

Interview Schedule #1 is designed for use with testing directors whose district has implemented or planned one or more parallel assessments. In addition to use and impact questions, it contains preliminary questions regarding which

of the four state test content areas the parallel assessment tests, what test specifications were used in test development, the respondent's rating of the degree of parallelness between the local parallel assessment and its corresponding state test, and information on grade levels of participating students and frequency and time of year of parallel assessment administrations. Responses to Interview Schedule #1 questions were required for each parallel assessment separately although recorded simultaneously to avoid repetition for testing directors with several parallel assessments to report. Interview Schedule #2 is designed for use with testing directors whose district has not implemented or planned a parallel assessment. In these interviews respondents are asked to answer the same use and impact questions, but for a hypothetical situation; that is, what purposes might be served and benefits and drawbacks might be expected if that respondent's district were to implement parallel assessments in two subject areas.

In both types of interviews, respondents answered questions about likely uses of scores from parallel assessments and likely impacts. Regarding uses of scores, interviewers posed the question "Does your LEA use scores from your parallel assessment in any of the following ways?" followed by seven potential score uses (which are excerpted in the body of Table 3 below). Respondents indicated whether each use existed in their districts by responding "Yes" or "No" and elaborating or explaining voluntarily or after probing. Next, interviewers posed the question "As a result of your LEA's parallel assessment activities, are any of the following actually or likely to be impacts?" This introductory question was followed by 16 potential impacts (excerpted in the body of Table 4 below) which respondents indicated had been, in their judgment, positive or negative impacts (i.e., benefits or drawbacks) or of no impact.

Analysis of Impacts on State Test Pass-Rates

Percentages of ninth-grade students in each school district who pass state functional tests the first time they take them and parallel assessment implementation dates were examined for evidence of impacts of local parallel assessments on performance on state competency tests.

Descriptions of Four Local Parallel Assessments

A Parallel Writing Assessment in Washington County

Context. Washington County is a predominantly rural area in western Maryland, situated between central Pennsylvania and West Virginia, with several pockets of industry. The Washington County school system consists of 24 elementary, seven middle, one middle-senior, and seven

high schools and approximately 17,000 students and 1,900 instructional and support staff. Student achievement on norm-referenced achievement tests is slightly below the state average. On state competency tests, Washington County 9th graders recently have performed just above the state pass-rate in reading and citizenship and just below the state pass-rate in mathematics and writing.

Purpose. The county writing assessment has two main purposes. First, an original purpose was to acquaint students with the format of the state writing test since direct writing assessments were a novel assessment procedure in the early 1980s. Second, to obtain diagnostic information on strengths and weaknesses of individual students and classes so that intervention can take place prior to administration of the state test in grade 9. This information is used by middle school language teachers to determine which aspects of their writing instruction have been successful and where to improve instruction.

Development and participation. Washington County's writing test is intended as an assessment parallel to the Maryland Writing Test and for use with all grade 7 students in the county. The county language arts supervisor directed the development of procedures, materials, and prompts for the first and subsequent writing tests following approval by the superintendent. County personnel wanted to develop a test for use with 7th graders in place of the state's multiple-choice writing test.

The county writing test was first administered in Spring 1983. Like its state counterpart, the Washington County parallel writing assessment requires a narrative and an explanatory essay. The format and content of the essay prompts, administration procedures, and scoring procedures and rubrics are closely modelled after state writing test procedures. The language arts supervisor leads prompt development each year with a committee of the seven middle school English department chairpersons. The department chairpersons have been trained in prompt development and essay scoring by staff of the state department of education. Using abbreviated test specifications in instructional guides published by the state, the committee develops three narrative and three explanatory prompts. These candidates are reviewed by all middle school English teachers, whose comments are then used by the prompt development committee to revise and select prompts for administration to students.

Each Spring, all grade 7 English teachers administer the parallel writing test during an extended, 90-minute English class to all grade 7 students. Student essays are scored during the summer by middle and high school English teachers who have been trained by a high school English department chairperson using the state's scoring rubric and

anchor sets developed by the language arts supervisor and scoring trainer. Training is completed in approximately two to three hours and scoring commences immediately. Although teachers are not required to pass qualifying rounds during training, inaccurate scorers are identified and assigned to non-scoring duties, and their scores are corrected. Participating teachers are paid standard county workshop fees. A plan to score and report the essays during the school year is under consideration.

At the opening of school in grade 8, students receive their essays, scores, and a written explanation of their scores and in what ways they can improve their writing. The essays and scores are filed in the student's permanent writing folder for use by grade 8 English teachers. The folder follows the student's other records through high school. Grade 8 and 9 English teacher compare the essays from the grade 7 parallel writing assessment to subsequent writing samples taken in grade 8 English classes, which are also placed in the permanent writing folder, to determine and document progress in writing. English department chairpersons and the county supervisor examine school and county mean scores from each year's parallel writing assessment to evaluate county writing instruction efforts.

Statistical characteristics. Some analysis of scores has been conducted to assess the validity of the county writing assessment for predicting performance on the state writing test. Table 1 below contains descriptive statistics for the county test.

Table 1

Descriptive Statistics for the Washington County Parallel Writing Assessment: Grade 9 Students in 1985-86 and 1986-87

Local Test	N	Mean (S.D.)[1]	Reliability[2]	Validity[3]
Writing, Grade 7				
1985-86	1,209	5.0 (1.4)	.71	.44
1986-87	1,205	5.3 (1.3)	.70	.48

1. Total essay score on a 0.0 to 8.0 scale.
2. Coefficient alpha calculated based on variance of narrative and explanatory essay scores separately and on total essay score variance.
3. Predictive validity coefficient is correlation between total holistic scores on local test and state tests administered during grade 9.

Uses and impacts. Students receive their essay scores and annotations at the beginning of grade 8 for use in determining their strengths and weaknesses as writers. In addition, the county superintendent, director of instruction, and language arts supervisor receive a report of average parallel assessment scores for each school and the county. Using this report and information on the relationship between the county and state tests (see table above), they develop a sense of the likely performance of these 8th grade students on the state's writing test (given in grade 9), and adjust and focus instructional plans.

The county parallel writing assessment has created several impacts which may not have occurred without the assessment, some which are benefits, some drawbacks. For example, students are quite familiar with the format of the state writing test by the time they take it in grade 9. Approximately 60 English teachers have participated in training and scored county writing assessments and use holistic scoring to evaluate classroom writing assignments. Their understanding of the requirements of the state test is "one hundred percent improved" according to the language arts supervisor, and this understanding helps teachers focus middle school writing instruction. English teachers are now using prompt formats for literature, book report, letter-writing, and other writing assignments. Teachers are careful to file student writing assignments and scores in the permanent writing folders to document instruction and show progress in writing skill. Finally, teachers in other subject areas such as science are giving more writing assignments and direct writing instruction is beginning in the elementary schools.

Two drawbacks to the county's parallel writing assessment should be noted. First, middle school teachers sometimes feel defeated if their students' scores on the county assessment are low. (Grade 9 teachers have responded similarly in the past to performance on the state test.) Also, some teachers feel that the increased focus on the county and state writing tests has replaced instructional time formerly used for creative writing assignments.

We believe that as a result of the parallel writing assessment teachers, students, and the community become aware of high school graduation requirements earlier than in the past. Middle school teachers are more aware of the state testing program and, therefore, there is a high degree of match between middle school writing curriculum and instruction and the writing skills required by the state test. Overall, the county parallel assessment has helped bring about more positive attitudes among teachers and students toward the state writing assessment program.

Citizenship and Writing Parallel Assessments in Charles County

Context. Charles County is a predominantly rural community in southern Maryland across the Potomac River from Virginia. The Charles County school system consists of 16 elementary, 6 middle, and 4 high schools and 17,000 students and 1,600 instructional and support staff. Student achievement on norm-referenced achievement tests has been slightly below the state average. On state competency tests, 9th graders recently have performed equal to the state pass-rate in reading, below in citizenship, and substantially above in mathematics and writing.

Purpose. The Charles County citizenship and writing parallel assessments have been designed for different purposes. The Charles County Interactive Testing for Citizenship (CCIT-Cit) is a computer-adaptive parallel form of the state test of citizenship. Its purpose is to help teachers assess each student's level of mastery in the three content subdomains that are tested on the state test. The Charles County Writing Assessment (CCWA) was designed to meet five objectives: (1) To provide students with an opportunity to take a test that closely resembles the state's writing test; (2) To provide students with information about their writing performance and ways in which it can be improved; (3) To provide teachers with training in holistic scoring methods; (4) To provide teachers with a standard indicator of students' writing; and (5) To provide schools and the district with information concerning the current status of writing performance.

Development and participation. The CCIT-Cit was developed during the Spring of 1987 and was first available for student and teacher use in the Fall of 1987. The technical development of this assessment (item evaluation, calibration, and equating, software adaptation, etc.) was carried out by the Office of Student and Program Evaluation; the development of the item pool was under the authority of the supervisor for social studies. The CCWA is given to students in the Spring of grades 5 (CCWA-5) and 8 (CCWA-8). A field test version of the CCWA-8 was administered during the 1984-85 school year; the CCWA-5 and CCWA-8 were first administered in the 1985-86 school year. Initial development of and improvements in the CCWA have been collaborative efforts between the county testing and language arts supervisors.

The CCIT-Cit. The CCIT-Cit was incrementally developed over a number of years. In the spring and summer of 1985, local social studies teachers developed a bank of 420 test items using available state specifications. During the 1986-87 school year the items were reviewed for technical quality and the bank was reduced to a set of 312 which were

jassembled into six spiraled test forms. These paper-pencil forms were administered to the entire ninth grade (N=1,262) such that no item was taken by fewer than 392 students. The forms were administered within two weeks before or after the administration of the state citizenship test and state and local items were calibrated simultaneously using the BICAL program. State item bank difficulty values were regressed on the observed across-form mean difficulties for the 45 state items that were administered concurrently with local items. The resulting regression equation was used to link local item difficulty values expressed on the state test's item difficulty scale. (The correlation between predicted and bank difficulty values was so high -- 0.96 -- that the linking was accomplished essentially by the regression intercept term.) This regression approach was used because of the relationship between the two tests to be equated; that is, the state test is a criterion measure whose scores are to be predicted with local test scores. The equated difficulties are used for converting student ability estimates to a state scale-score metric.

Each of the local items was entered into item banks in the MicroCat computer-adaptive software package. There are three banks, one for each of the content subdomains on the test. The software is programmed to provide students with a point estimate and a 90 percent confidence interval estimate of their scores. Scores are stored on disk along with the student's response record so that the teacher can review a student's performance from any particular test administration. The CCIT-Cit was installed at the two high schools in the county that currently have computer labs capable of handling the computer-adaptive testing software. Students use the labs at any scheduled time. Typically, students use CCIT-Cit two to three times prior to their taking the state citizenship. Test administration takes roughly 15 minutes. End-of-test criteria are 20 items or an ability estimate standard error of 0.20, whichever comes first.

The CCWA-5 and CCWA-8. Close scrutiny has been paid to assure that both CCWA-5 and CCWA-8 match the state writing test to the extent that local resources allow. The development of the prompts and scoring rubric are carried out by the local testing supervisor, the local language arts supervisor, and the county's four writing resource teachers. State guidelines for prompt and rubric development and for test administration are consistently followed in the local assessment.

The CCWAs depart to some degree from the state in terms of the methods used for scoring and score reporting. Because less time and money are available for training readers, it is more difficult to assure that readers score reliably and to compensate for or correct unreliable

readers. Estimates of the correlation between first and second ratings of each essay for the 1987 CCWAs are between .70 and .75. Given local resource limitations, these correlations compare favorably with those obtained on the state writing test (.89 and .92 in 1987). Local score reporting attempts to minimize the stakes associated with CCWA scores: no information on the CCWA score is maintained in a student's permanent record, and scores are printed on the student's writing sample in order to emphasize that the CCWA primarily is a means of acquiring a sample of writing, rather than a means of obtaining a score.

Uses and impacts. Students and teachers use CCIT-Cit scores primarily as predictors of future test performance. When CCIT-Cit performance indicates that a student is below criterion, instructional activities are provided to increase the student's knowledge in that domain. Teachers and students are provided with estimated scale scores and 90 percent confidence intervals around that scale score. These data are stored on disk; as students finish testing, their estimated likelihood of passing that domain on the state test is displayed as one of five levels: 20 percent or less chance of passing, 20-40 percent chance, 40-60 percent chance, 60-80 percent chance, and a greater than 80 percent chance of passing. These probabilities are based upon the error variance criterion used as a stopping rule in the computer-adaptive testing procedure. The decision to report these probabilities was made to compensate for any scaling and prediction inaccuracies.

It is too early to assess the direct benefits of the CCIT-Cit. Some preliminary observations, however, appear warranted. For the fall, 1987 administration of the state citizenship test, the two CCIT-Cit schools had a combined pass rate of 48.6 percent for 10th grade retest students compared with a combined pass rate of 26.8 percent for the two non-CCIT-Cit schools. A secondary benefit is that MicroCAT can be used to print paper-and-pencil tests according to user-defined specifications. Teachers at the two non-CCIT-Cit schools have used this capability and report satisfaction with the quality of the tests it produces.

Performance on the CCWAs is reported and used in a much different fashion from CCIT-Cit performance. CCWA performance is exhibited both as a score and as the writing sample itself. It is typical for teachers to use the writing sample itself rather than the writing score for instructional purposes. Using CCWA results for class placement is discouraged because of concerns about score reliability and county philosophy that important decisions like class placement should not be made based on such a narrowly focused assessment. Schools use school and classroom score distribution information to assess the

progress of the school writing program. Aggregate units are encouraged to compare themselves with the next-highest aggregate. That is, classrooms compare themselves to the school, schools to the county. Because there is no higher aggregate for the county, we do not use score distributions for county-wide evaluation.

A number of benefits have accrued from the CCWAs, some unanticipated at the time of development. The following comments represent a review of comments made by teachers who have participated in CCWA scoring workshops. For the most part, the benefits fall into two major categories: (1) Improved attitudes toward efforts to assess writing; and (2) Improved delivery of writing instruction. Teachers who have participated in CCWA holistic scoring workshops report that their participation has helped them become more appreciative of standardized direct writing assessments. Improvements in instruction have, by and large, also been a direct result of teachers' being involved in the CCWA scoring workshops.

Teachers point to three topics of improvement: clarified expectations, improved between-school communication, and expanded awareness of student performance. By "clarified expectations" teachers mean that their first-hand experience with the CCWA has helped them form clearer operational definitions of state scoring criteria such as elaboration, sufficient detail, and sequence. "Improved between-school communication" has been an unexpected benefit of the CCWA. Isolation is often an overlooked aspect of teacher life. Attending a CCWA scoring workshop provides teachers with the opportunity to participate with colleagues in a professional activity that has direct bearing on their delivery of instruction. "Expanded awareness of student performance" is a corollary to the communications issue. Here, teachers report benefits not from being able to talk to other teachers but from being able to see the work of other students. Teachers report being able to carry back to their schools a more accurate perception of the distribution of writing performance across the county. A fourth benefit of CCWA to instruction is not reported by participants in the CCWA workshops. The four writing resource teachers report that CCWA-8 has been a significant factor in preparing for incoming 9th grade students. The writing resource teachers are able to review the performance of the incoming class for their high schools and identify students with particular needs. The student's CCWA-8 document is also used for re-writing exercises at the beginning of grade 9. This eliminates the week it typically takes to assign, receive and score a first piece of writing from students.

Mathematics, Citizenship, and Writing Parallel Assessments in Cecil County

Context. Cecil County is a largely rural area in the northeast corner of Maryland, about mid-way between Baltimore and Philadelphia. The Cecil County school system consists of 16 elementary, four middle, two junior-senior high, and three high schools and 12,000 students, 700 teachers, and 400 support personnel. Student achievement on norm-referenced tests is approximately at the state average. However, student performance on the four state competency tests has been consistently above the performance of the rest of the state.

Purpose. As state-mandated graduation tests were implemented in the early 1980s, subject area supervisors and the testing director in Cecil County began devising strategies to prepare students for their first encounter with the tests and for subsequent attempts if they failed. Supervisors of mathematics, language arts, and social studies decided that locally developed parallel assessments would be effective tools.

Each supervisor worked independently, but all agreed that one basic criterion was to match the state tests as closely as possible. The supervisors wanted to develop tests to gather information to be used to guide remediation, to help place students in appropriate instructional groups, to assess curriculum strengths and weaknesses, to serve as practice tests for the state tests, and for use in staff development. The mathematics, citizenship, and writing parallel assessments were fully developed in 1983, 1984, and 1985 respectively.

Development and participation. Each of the three tests was developed independently and implemented at different times between 1983 and 1985. The content area supervisors and local testing director led item development workshops in which special basic skills remediation teachers wrote and reviewed test items. Abbreviated test specifications for each state test, published by the state department of education, guided item development.

The mathematics and citizenship tests. The mathematics and citizenship skills tests, like their state counterparts, consist of multiple-choice items that assess specified objectives within specified content domains. Large pools of items were developed to cover all objectives assessed by each test. Item format and contents were reviewed by state department content area instructional specialists to improve parallel construction. Each test contains 100 items to facilitate score interpretation. Each year the tests are reviewed and revised using results from the previous year's

administration. Reviews have led to addition of items that assess the more difficult objectives and reduction and deletion of items that assess easier objectives.

At the end of the school year all grade 7 and 8 students take the parallel mathematics test in preparation for the state test administration in fall of grade 9. All 8th graders take a one-domain form of the parallel citizenship test after they complete an instructional unit on the U.S. constitution. All 9th graders take a longer parallel citizenship test covering all content domains approximately one month prior to administration of the state test. The local tests are scanned in the district central office and results are returned directly to each school. Mathematics results are available at the opening of school for 8th and 9th graders; citizenship tests are scored as quickly as possible to guide preparation for the state test administration. Number and percent-correct scores for individual students, and average scores for each school and the entire county are reported for each test objective and for the total test, and schools are ranked according to their average total test scores.

Writing test. Like the state test, the local parallel writing test requires students to write an essay on each of two days to a narrative and an explanatory prompt. Students follow state administration procedures on the local writing test: they write a draft, revise the draft using a checklist, and write final copies on separate paper. All 7th graders take the parallel writing test in the spring; individual, school, and district narrative and explanatory essay scores are available for the opening of school in grade 8.

A key component of any direct writing assessment is scoring. The local writing test is scored using locally developed training materials, and scoring rubrics and procedures developed for the state writing test. Scorer training is conducted during the school year by the language arts supervisor and is limited to one day. Approximately 1,000 narrative essays are scored twice each over a three-day period, and the cycle is repeated for explanatory essays. Scoring is conducted by a core of special writing assistance teachers, but at least half of the scorers are volunteer English and special education teachers and occasional administrators.

Statistical characteristics. Some analysis of scores has been conducted to assess the validity of Cecil County tests for predicting performance on state tests. Table 2 below contains descriptive statistics for the three local tests.

Table 2

Descriptive Statistics for Parallel Assessments in Cecil County: Grade 9 Students in 1986-87

Local Test	N	Mean (S.D.)[1]	Reliability[2]	Validity[3]
Citizenship				
Grade 9	805	67.0 (17.5)	NA	.79
Mathematics				
Grade 7	737	74.7 (17.8)	NA	.76
Grade 8	684	83.0 (15.1)	NA	.82
Writing				
Grade 7	680	5.6 (1.20)	.70	.50

1. Percent correct scores for citizenship and mathematics tests; for writing, total essay score on a 0.0 to 8.0 scale.
2. Coefficient alpha calculated from variances of narrative and explanatory essay scores separately and from total essay score variance.
3. Predictive validity coefficient for citizenship and mathematics is correlation between percent correct scores on local test and state tests administered in grade 9; for writing is correlation between total holistic essay scores on local test and on state test administered in grade 9.

NA: Not available at this time.

Uses and impacts. The county content area instructional supervisors perceive both benefits and drawbacks associated with their parallel assessment programs, but on the whole believe that the benefits outweigh drawbacks. They report that, in general, parallel assessments provide teachers with the information they need to deliver focused preparatory and remedial instruction but represent an additional intrusion on instructional time. (However, they do recognize in other contexts that parallel assessment information makes instruction focused on state-mandated curriculum more efficient.)

One of the major benefits of all three parallel assessments is the increase in teacher knowledge of the content, format, and scoring and interpretation of state competency tests. This increased knowledge and awareness allows teachers to target instruction on skills and content contained in state tests and to emphasize specific test objectives and content area domains for students who perform poorly on local parallel assessments. In addition, students who perform well on local assessments are able to spend less instructional time on preparation and remediation because

parallel assessment results are used to place students in instructional groups. Parallel assessments have provided practice in dealing with formats and content of the state tests which are unfamiliar to students, especially in the citizenship and writing areas. Supervisors have also used parallel assessment results to refine the curriculum restructuring that took place when state tests were first implemented.

The state competency testing program has had a significant impact on county mathematics, social studies, and writing instructional programs. The Cecil County parallel assessments have contributed to that impact and the impact has been generally positive. Social studies instructional staff are unhappy, though, with they way in which instructional focus has shifted toward citizenship instructional objectives and away from other social studies goals. However, all content area instructional supervisors believe that teachers have a greater knowledge and appreciation of the basic skills instruction and assessment program for which they are responsible. Students also know early in their secondary programs what is required of them for graduation.

A Parallel Computer-Adaptive Mathematics Assessment in Montgomery County

Context. Montgomery County is a predominantly suburban area with large concentrations of high technology businesses. Since it is adjacent to Washington, DC many residents are professionals who work in downtown Washington. A large area of the northern section of the county is rural but undergoing rapid residential development. The Montgomery school system consists of 90 elementary, 21 middle and junior high, and 37 high schools and approximately 94,000 students and 12,000 instructional and support staff. It is one of the largest and wealthiest school systems on the East Coast. Student performance on norm-referenced achievement tests has been well above state averages. County pass-rates on all state competency tests have been among the highest of all school districts in the state.

Purpose. In Spring 1987 the Montgomery County school system implemented a local parallel assessment known as the Computerized Adaptive Functional Mathematics Assessment (CAFMA). CAFMA is an outgrowth of Project Adapt, a collaboration between the Montgomery County Public Schools and Maryland State Department of Education, in which state competency tests in reading and mathematics were pilot tested as computer-adaptive tests. Results of two years of pilot testing have been reported elsewhere (Salehi & Stevenson, 1987). The computer-adaptive reading and

mathematics tests have proven to be highly efficient and reliable alternatives to the state paper-pencil competency tests. The impetus for the CAFMA comes from the Montgomery County Department of Educational Accountability, and in light of the positive results from Project Adapt. The purpose of the CAFMA is to provide math teachers with an instrument that can efficiently assess a student's likelihood of passing the state math competency test.

Development and participation. Montgomery schools testing staff developed the CAFMA item bank using items from the state's grade 7 mathematics practice test. These 77 non-secure items are part of the state's larger, secure item bank which is used for constructing annual editions of the state mathematics competency test. Since the state practice test items were generated from state test item specifications and are linked to the state item bank using Rasch equating of item difficulties, the CAFMA is a parallel assessment of the state graduation test and an equivalent form of the state practice test.

The CAFMA is administered on IBM-workalike microcomputers by MicroCat software. Items are presented on the screen and students enter responses using the keyboard. Instruction on using the keyboard and how to respond to items is presented on the computer. Testing begins with administration of an item of average difficulty which the computer randomly selects from the item bank. The computer scores the student's response, estimates the student's ability, and selects the next item following typical computer-adaptive testing procedures. A Bayesian strategy guides item selection at every stage of the adaptive-testing process, while a maximum-likelihood estimation procedure estimates student ability. In a CAFMA domain procedure testing ends after 10 items from one content domain have been administered. In a CAFMA across-domain procedure testing ends after administration of 25 items or when the ability estimate standard error reaches 0.10. Help screens guide teachers in setting up testing sessions and in displaying, printing, and saving test results. On the average, students complete domain tests in about 12 minutes and across-domain tests in about 25 minutes.

Students in grades 7-12 take CAFMA tests after completing units of focused preparatory and remedial instruction. Previous research has indicated that the CAFMA has no practice effects associated with it (Salehi & Stevenson, 1987). It is administered periodically and at the discretion of the individual teacher to students in high school math labs after they have participated in preparatory or remedial instruction and as soon as feel they have mastered the skills required to pass. Performance on the CAFMA is reported in the form of theta estimates and thetas transformed to state test scale scores. Decisions on how to

use information from CAFMA administrations are left to the individual teacher. As a result of teacher comments, the CAFMA "passing score" (350 on an approximate 200-450 scale) is 10 points higher than the state passing score. This is intended to minimize false positive mastery classifications and to avoid falsely raising student and teacher expectations.

Uses and impacts. Teachers use the CAFMA primarily as a screening instrument: that is, to indicate whether a student is likely to pass the state competency test. Some teachers also use the CAFMA to assess the likelihood that a new transfer student will pass an imminent administration of the state test. If students "fail" the CAFMA they continue to receive additional instruction in the functional mathematics skills assessed on the state test. Since the CAFMA can also provide "pass/fail" information at the content domain level, teachers also use scores from domain subtests to identify weaknesses and target instruction. CAFMA scores from domain subtests are valuable in helping channel efforts to improve school and county pass-rates on the state mathematics competency test. No evidence has yet been collected, however, on the extent to which CAFMA can help improve student performance on other standardized tests or on classroom assessments. As a result of teachers' comments about the CAFMA, a prototype procedure is under development to link CAFMA diagnostic information to a directory of instructional materials and activities using a commercially available artificial intelligence software program.

Several benefits have become evident to teachers who have used the CAFMA. Foremost among these is that student attitudes toward the state mathematics competency test become more positive. In addition, and because of the ability to assess students at the content domain level, both teachers and students are more aware of the mathematics skill requirements of the state test. This awareness has not led to "teaching to the test," but rather to identification of appropriate instructional materials that will help students master required skills and concepts. Thus far, use of the CAFMA has had no impact on county mathematics curriculum nor resulted in increased focus on the state competency testing program in staff development activities.

Telephone Survey of District Testing Directors: Results and Discussion

The 20 school district testing directors not involved in conducting this study were interviewed to determine the numbers, features, purposes, uses, and impacts of local parallel assessments. The information from these interviews

indicates the extent and nature of parallel assessment activities in Maryland beyond the four district programs described above. Of the 18 districts interviewed thus far, 12 report having implemented or planned 31 separate parallel assessments. Of these, eight assess reading, nine assess mathematics, five assess citizenship, and nine are parallel writing assessments. The reading and mathematics parallel assessments are particularly noteworthy because the state department of education offers optional grade 7 diagnostic reading and mathematics tests that produce scores that are equated to the grade 9 graduation test scores.

Characteristics of the Parallel Assessments

District testing directors report that the abbreviated item specifications available in instructional guides published by the state are the sources for item generation for 29 of the local tests. Of the remaining two, items for a reading assessment were generated from corresponding curriculum objectives prior to development of the state reading test, and one source is unknown. The 12 directors report that, in general, these local assessments are "somewhat" to "very parallel" to their state test counterpart content and procedures (five-point scale ranging from "not at all" to "very parallel": Mean=4.39, S.D.=0.56). Most local tests are administered to students in grades 7 through 9 prior to state test administration and to students in grades 9-12 who fail state tests. Twenty-four of the assessments are administered only once per year, three twice per year, and the rest several times per year. Parallel assessment administration dates coincide with state test administration dates and preparatory instruction.

Uses of Scores from Parallel Assessments

Table 3 below summarizes responses of the 12 testing directors in districts with parallel assessments to seven interview items regarding the question "Does your LEA use scores from your parallel assessment activities in any of the following ways?" Numbers and percents of responses for each potential score use are reported across the 31 parallel assessments and for each test content area separately. Since the number of respondents is fairly low, differences between percentages should be interpreted cautiously (see Table 3).

Table 3 here

As Table 3 indicates, LEA staff use scores from local parallel assessments in the ways they are most useful. In general, these uses are consistent with the score uses reported in the four detailed descriptions in the previous

section of this paper. However, on almost no score use was there 100 percent agreement. The 12 responding testing directors report that, in general, parallel assessment score information is used for (1) Placing students in instructional groups; (2) Selecting instructional activities; (3) Documenting delivery of preparatory and remedial instruction; and (4) Improving performance on corresponding state competency tests. The four negative responses regarding placement of students are consistent with reservations expressed in one county parallel assessment description regarding student placement decisions based on a single, narrowly focused assessment. Also, the six negative responses regarding documentation of service delivery probably reflect the existing documentation resulting from placement of students in specific courses. Respondents report that scores from parallel assessments are also used to improve teacher classroom instruction and assessment activities. These 26 positive responses probably indicate in which LEAs teachers have reported or demonstrated that their experience in developing items, scoring, and interpreting scores from parallel assessments has enabled them to focus their instruction on skills and information assessed on state competency tests. In addition, their responses echo comments about professional development reported in the four previous LEA descriptions.

Respondents are almost unanimous in indicating that their districts do not use scores from parallel assessments to aid improvement of student performance on other standardized tests. However, they are divided on the use of scores for helping improve student performance in the classroom. These results probably indicate the perceived limited overlap or commonality between state basic skills requirements, as indicated by state competency test content, and content of other achievement tests and, especially, what teachers teach and test in the classroom.

Finally, no striking differences are apparent in score uses for the four test content areas, except perhaps for the use of reading parallel assessment scores for improving teacher and student classroom performance. In this case, it appears that reading scores are more useful for helping students than scores from the other areas, but less useful for helping teachers improve their own classroom performance. This is not a surprising disparity given grade 9 state pass-rates on the reading test (consistently around 93 percent) and on the other three state tests (approximately 67 percent in the last two years). In this context the disparity in parallel assessment score use might suggest that local reading tests help identify reading deficiencies of secondary level students who experience chronic reading problems on which teachers can focus instruction; but little help in actually remediating these chronic difficulties.

Impacts of Parallel Assessments

Table 4 below summarizes responses of the 12 testing directors to questions regarding perceived impacts of parallel assessments on attitudes toward the state testing program and on local curriculum, instruction, assessment, and staff development activities. Numbers and percents of positive, negative, and non-existent impacts are reported across the 31 parallel assessments and for each test content area separately. Since the number of respondents is fairly low, differences between percentages should be interpreted cautiously (see Table 4).

Table 4 here

Table 4 indicates that, in the view of the 12 testing directors, parallel assessment activities have helped to engender improved attitudes toward the state competency testing program and increased (1) Focus on state basic skills objectives and state competency test content; (2) Early awareness of graduation test requirements; and (3) Focus on state tests in staff development activities. Responses in Table 4 also indicate that there has been either no impact or a positive impact on community attitudes toward the state competency testing program, on awareness of basic skills instructional program deficiencies, and on the quality of staff development activities. In contrast, the impact on curriculum has been perceived generally as either negative or non-existent. That is, responses suggest that parallel assessments have contributed to narrowing and shifting of curriculum that may have occurred with the introduction of high-stakes tests competency tests, but that little expansion of curriculum has occurred as a result of parallel assessments.

Although views are mixed, in general the 12 respondents report little impact of parallel assessments on classroom assessment activities (impacts 13 and 14). Responses are also split on impacts on "teaching to the (state) test" and away from teaching mastery of skills and concepts. (In the interviews we defined "teaching to the test" as direct teaching of simulated or actual state competency test items or the use of frequent practice testing with little intervening instruction.) Although a majority of the responses indicate that parallel assessments have not increased this inappropriate use of tests, respondents reporting this impact see it both as a positive and negative impact (see impact 9). The seven reported instances of negative impact may suggest cases where teaching to the test has occurred; the five instances of positive impact may indicate misinterpretation of the question.

Some responses to questions about impacts of parallel assessments differ by test content area. For example, responses and probed elaboration of responses suggest that improvement in attitudes toward the state competency testing program occurred mostly due to parallel assessment activities in mathematics and writing. The state mathematics and writing tests reportedly are sources of great anxiety for students; parallel assessments in these areas may reduce this anxiety and improve attitudes by providing familiarity with test format, content, and general difficulty level. Also, responses at impact 10 suggest that writing and reading parallel assessments have contributed to the narrowing of English and writing curriculum. The responses regarding writing parallel assessments are consistent with complaints during the interviews and in the four previous detailed descriptions that focus on state writing competencies has eroded instructional time available for creative writing assignments. Finally, responses regarding the impact of citizenship parallel assessments on formats for classroom and homework assignments are consistent with public perceptions of the state test: that it measures "trivial" facts rather than skills and knowledge necessary to function as a citizen and voter.

Results from Interview Schedule #2

Six of the 18 district testing directors report no parallel assessment activities in their districts. We asked these directors the same use and impact questions as with directors in other districts, but for a hypothetical situation in which their district has implemented two parallel assessments. Their responses to the interview questions indicate their views on the potential uses of scores and potential positive and negative impacts from parallel assessments. Since the number of respondents is low, responses should be interpreted cautiously.

Responses of this group regarding potential uses of scores from parallel assessments are similar to responses regarding actual uses, except for three uses that would become clear probably only after implementing parallel assessments. Unlike the other group of testing directors, these respondents are split regarding using parallel assessment scores for student placement (3 yes, 3 no) and to improve classroom instruction and assessment activities (3 yes, 3 no). While the group of testing directors in districts with parallel assessments are unanimous on using scores from parallel assessments to improve state test pass-rates, this group of respondents has yet to be clearly convinced of this potential use (4 yes, 2 no).

As with potential uses, the testing directors in this group report expectations for impacts of parallel assessment activities that differ from actual reported impacts. For example, these respondents agree with the other group regarding positive impact on students' and administrators' attitudes toward the state competency testing program (5 expect positive impact) but are not aware of the lack of impact also reported (1 expects no impact on students, none on administrators). These responses appear to suggest that testing directors in districts with no parallel assessments are more optimistic about a positive impact on their own attitudes toward the state testing program than their colleagues actually experience! Also, these respondents are not aware of the reported lack of (1) Impact on narrowing of curriculum; (2) Increases in "teaching to tests"; and (3) Increases in classroom testing of recall over comprehension. Further, they are unaware of the reported benefits to the match between instructional content and state test content and to the quality of staff development activities.

Impacts of Parallel Assessments on District Pass-Rates

The most widely cited and closely watched indicator of performance on Maryland competency tests is the grade 9, first-time-taker pass-rate. This percentage of students who pass each test is announced with other competency test performance information by the state superintendent of schools. District superintendents also announce their district pass-rates. Generally speaking, grade 9 pass-rates indicate the "success" of kindergarten through grade 8 basic skills instructional programs and suggest the magnitude of the task of remediating basic skills of high school students prior to graduation.

Given the apparent power of these indicators, it seems reasonable to expect that local testing directors and instructional supervisors might evaluate the validity and usefulness of their local parallel assessments by examining the impact using the tests on grade 9 pass-rates. In addition, it seems reasonable to expect individual districts' grade 9 pass-rates to increase after cohorts of students take parallel assessments and participate in instruction targeted by this information. Simply put, it seems that some clear benefit to performance on state tests would have to be demonstrated to warrant the expense and efforts required to develop, administer, score, and report parallel assessments.

To investigate the impact of parallel assessment activities on student performance on state competency tests, we examined grade 9 pass-rates since 1984 for each district on all four state competency tests. No simple picture

emerges from this examination. Several districts with parallel assessments have shown a steady increase in pass-rates over the last three to four years. However, as illustrated by excerpts from several years' pass-rates below, districts with parallel assessments show both pass-rate increases and decreases while state pass-rates remain stable (reading) or increase each year (mathematics, citizenship, writing). Also, some districts with no parallel assessments have shown substantial pass-rate increases (see Table 5 below).

Table 5

Selected District Pass-Rates on State Competency Tests, 1984-1987

	1984	1985	1986	1987	Type of pass-rate change
Districts with parallel assessments					
Parallel assessment implemented in:					
Reading (Fall 1982)	96.7	94.7	94.9	95.4	Little change
Mathematics (Spring 1985)	68.7	62.3	58.5	63.1	Up and down
Citizenship (Spring 1985)	19.6	71.2	65.8	78.2	Up and down
Writing (Spring 1983)	47.3	43.4	66.5	61.2	Up and down
Districts with no parallel assessments					
Reading	91.0	92.5	90.5	93.3	Up and down
Mathematics	65.0	50.5	52.2	61.7	Up and down
Citizenship	34.5	48.6	58.9	72.0	Increase
Writing	56.5	68.4	78.8	73.3	Up and down
All districts combined					
Reading (Fall 1982)	93.1	92.2	92.2	93.1	Little change
Mathematics (Fall 1982)	61.0	64.9	65.5	68.3	Increase
Citizenship (Spring 1984)	42.1	58.6	66.4	73.3	Increase
Writing (Spring 1983)	51.4	54.1	68.5	67.1	Increase

Several explanations are possible for this apparent lack of relationship between parallel assessment activities and state test pass-rates. One such explanation is that positive effects of using parallel assessment information can be masked by a particularly poor or excellent performance of a single teacher or school. For example, in

a district with 900 grade 9 students, a single teacher who teaches 150 students each day could adversely affect the district's pass-rate by failing to prepare students for a state test administration. In contrast, the pass-rate of a large district with no parallel assessment could be increased by a singularly excellent performance of one or two schools. Individual school pass-rates have been shown over the years to fluctuate up and down even while district pass-rates follow a smooth trend. A second explanation -- and one that we as educators probably would prefer -- is that state competency test pass-rates change as a result of instructional efforts. Parallel assessments can only support and enhance the impact of instruction on test performance, not outstrip instruction's impact.

Summary and Conclusions

We began this paper by pointing out that the substantial literature on effects of state-mandated, minimum-competency high school graduation tests has very little to say about related local assessment activities. In this paper we set out to examine assessment-type responses to mandated state competency tests. First, we labelled local tests that are developed using state test specifications as "parallel assessments." Four Maryland school district testing staff then described in some detail well-established parallel assessment programs in their districts; they also conducted interviews with all other testing directors in the state. The interviews focused on uses of scores and impacts of parallel assessments. Finally, we searched for evidence of one expected impact: an increase in districts' state test pass-rates. We were able to describe several benefits and drawbacks to parallel assessment activities from the four detailed descriptions and from 18 interviews. Our current analysis of state test pass-rates did not clearly indicate a consistent positive impact on district pass-rates. Perhaps the most surprising finding from this study is the <u>extent</u> of parallel assessment activities: 31 separate tests in 12 districts (plus seven separate tests in the other four districts), some administered several times each year, all designed to aid student performance on state competency tests.

Several benefits and drawbacks are worth reiteration. First and foremost, teachers and instructional supervisors report with enthusiasm the benefits of participating in item development and essay scoring. These activities appear to aid their efforts in targeting instruction for particular students and groups and perhaps to improve classroom assessment strategies, especially in writing. Also, parallel assessments appear to improve administrators', teachers', and students' attitudes toward the state competency program by increasing familiarity with the

challenge they pose and by relieving anxiety. There is a cost, however, to enjoying these benefits. Responses indicate that the increased focus on basic skills objectives to the detriment of other learning outcomes is exacerbated by parallel assessments. A related drawback is the additional loss of instructional time necessary to administer parallel assessments. Interview responses also suggest that there may be some risk in parallel assessment activities of engendering inappropriate classroom teaching and testing practices such as teaching to tests, testing recall over comprehension, and using selected and short responses over constructed and extended response formats. Finally, several potential benefits to parallel assessments appear not to have been achieved. In addition to the lack of consistent impact on district pass-rates, these include transfer of improved performance to other assessment contexts such as norm-referenced achievement tests and informal classroom evaluation; more positive community attitude toward the state testing program; and increased awareness of instructional program deficiencies, which are reported to have been achieved only in some districts. In the final analysis, however, it appears that district testing and instructional staff feel that benefits from parallel assessments outweigh their drawbacks and time and money costs, especially in those districts where pass-rate increases have been small or inconsistent.

The results of this study raise several questions for the staff of Maryland's statewide testing program. State testing program staff have trained local staff in holistic essay scoring and provided some data analysis to support local parallel assessments -- should this support be extended to other areas? For example, should the state support, or even encourage parallel assessment activities by contributing and reviewing items? Or should the state department avoid involvement in local parallel assessment activities and even discourage such activities? Finally, do the benefits of these activities -- presumably, improved instructional and testing practice and increased learning and achievement -- outweigh the risks of abuse that could undermine state test validity?

Consideration of similar questions should also be of interest to other states with competency testing programs and to other high-stakes testing programs. For example, what is the extent of parallel assessment activities in other states? Are these activities supported, tolerated, discouraged, or forbidden? Will "raising the stakes" on the National Assessment of Educational Progress (NAEP) -- by implementing state-level sampling and interstate comparisons -- spawn parallel assessments? Are there locally developed assessments parallel to commercially produced norm-referenced achievement tests? How parallel are the practice tests and items used in review and preparation courses for

the various national admissions testing programs such as the PSAT, SAT, GRE, LSAT, and so forth? These are important questions to be answered in Maryland and elsewhere.

References

Anderson, B. L., & Hough, P. G. (1986, April). Do state testing programs constructively shape local curricula? Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Corbett, H. D., & Wilson, B. L. (1987, March). Local effects of state testing programs: Preliminary data from a two-state study. (Available from Research for Better Schools, 444 North Third Street, Philadelphia, PA 19123.)

Ferrara, S. F., & Thornton, S. J. (in press). Using NAEP for interstate comparisons: The beginnings of a "national achievement test" and "national curriculum." Educational Evaluation and Policy Analysis.

Popham, W. J. (1987). The merits of measurement-driven instruction. Phi Delta Kappan, 9, 679-682.

Salehi, S., & Stevenson, J. (1987, April). Project Adapt: An investigation of computerized testing procedures in comparison with paper and pencil testing procedures. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

sf1101

Table 3

Reported Uses of Scores from Local Parallel Assessments: 12 of 18 Testing Directors Reporting on 31 Parallel Assessments (in Numbers and Percents)

	All Tests		Reading		Mathematics		Citizenship		Writing	
	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
1. Place students in instructional groups.	27	4	7	1	8	1	4	1	8	1
	.87	.13	.88	.13	.89	.11	.80	.20	.89	.11
2. Select instructional activities.	31	0	8	0	9	0	5	0	9	0
	100	.00	100	.00	100	.00	100	.00	100	.00
3. Document delivery of preparatory and remedial instruction.	23	6	5	3	7	1	5	0	6	2
	.74	.19	.63	.38	.78	.11	100	.00	.67	.22
4. Improve performance (pass-rates) on state graduation tests.	31	0	8	0	9	0	5	0	9	0
	100	.00	100	.00	100	.00	100	0	100	.00
5. Improve student performance on other tests.	2	29	1	7	1	8	0	5	0	9
	.07	.94	.13	.88	.11	.89	.00	100	.00	100
6. Improve student performance in the classroom.	18	13	6	2	5	4	2	3	5	4
	.58	.42	.75	.25	.56	.44	.40	.60	.56	.44
7. Improve teachers' classroom instruction and assessment activities.	26	5	5	3	8	1	4	1	9	0
	.84	.16	.63	.38	.89	.11	.80	.20	100	.00
Total number of tests	31		8		9		5		9	

Note: Percentages may sum to less than 100 due to omitted responses or to more than 100 due to rounding.

Table 4

Reported Impacts of Local Parallel Assessments: 12 of 18 Testing Directors Reporting on 31 Parallel Tests (in Numbers and Percents)

Likely impacts are:	The impact is: All Tests +	−	0	Reading +	−	0	Mathematics +	−	0	Citizenship +	−	0	Writing +	−	0
1. Change in teacher attitude toward the state testing program.	19	3	9	4	0	4	6	0	3	2	2	1	7	1	1
	.61	.10	.29	.50	.00	.50	.67	.00	.33	.40	.40	.20	.78	.11	.11
2. Change in student attitude toward the state testing program.	19	4	8	4	0	4	6	0	3	3	2	0	6	2	1
	.61	.13	.26	.50	.00	.50	.67	.00	.33	.60	.40	.00	.67	.22	.11
3. Change in administrator attitude toward the state testing program.	21	0	10	4	0	4	6	0	3	4	0	1	7	0	2
	.68	.00	.32	.50	.00	.50	.67	.00	.33	.80	.00	.20	.78	.00	.22
4. Change in community attitudes toward the state testing program.	14	0	17	3	0	5	3	0	6	3	0	2	5	0	4
	.45	.00	.55	.38	.00	.63	.33	.00	.67	.60	.00	.40	.56	.00	.44
5. Increased focus on state basic skills objectives.	30	0	1	8	0	0	9	0	0	5	0	0	8	0	1
	.97	.00	.03	100	.00	.00	100	.00	.00	100	.00	.00	.89	.00	.11
6. Increased match between local instructional content and state test content.	31	0	0	8	0	0	9	0	0	5	0	0	9	0	0
	100	.00	.00	100	.00	.00	100	.00	.00	100	.00	.00	100	.00	.00
7. Increased awareness of instructional program deficiencies.	19	0	12	5	0	3	5	0	4	3	0	2	6	0	3
	.61	.00	.39	.63	.00	.38	.56	.00	.44	.60	.00	.40	.67	.00	.33
8. Early awareness of graduation test requirement.	26	0	3	6	0	2	7	0	1	5	0	0	8	0	0
	.84	.00	.10	.75	.00	.25	.78	.00	.11	100	.00	.00	.89	.00	.00
9. Increase in "teaching to tests" and away from skills and concepts.	5	7	19	1	2	5	2	1	6	1	1	3	1	3	5
	.16	.23	.61	.13	.25	.63	.22	.11	.67	.20	.20	.60	.11	.33	.56
10. Curriculum narrowing.	1	17	13	0	5	3	0	3	6	1	2	2	0	7	2
	.03	.55	.42	.00	.63	.38	.00	.33	.67	.20	.40	.40	.00	.78	.22
11. Curriculum shifting.	3	13	15	0	3	5	2	3	4	1	2	2	0	5	4
	.10	.42	.48	.00	.38	.63	.22	.33	.44	.20	.40	.40	.00	.56	.44
12. Curriculum expansion.	2	6	23	0	1	7	0	1	8	1	2	2	1	2	6
	.07	.19	.74	.00	.13	.88	.00	.11	.89	.20	.20	.40	.11	.22	.67
13. Increased classroom assessment of recall over understanding.	4	4	23	1	0	7	2	1	6	0	3	2	1	0	8
	.13	.13	.74	.13	.00	.88	.22	.11	.67	.00	.60	.40	.11	.00	.89
14. Increased use of short answers and selected responses in classroom and homework assignments.	8	2	21	1	1	6	2	0	7	4	1	0	1	0	8
	.26	.07	.68	.13	.13	.75	.22	.00	.78	.80	.20	.00	.11	.00	.89
15. Increased focus on state tests in staff development.	23	1	7	5	1	2	6	0	3	3	0	2	9	0	0
	.74	.03	.23	.63	.13	.25	.67	.00	.33	.60	.00	.40	100	.00	.00
16. Improvement in quality of staff development.	15	0	14	3	0	5	4	0	4	3	0	2	5	0	3
	.48	.00	.45	.38	.00	.63	.44	.00	.44	.60	.00	.40	.56	.00	.33
Total number of tests	31			8			9			5			9		

Note 1: Percentages may sum to less than 100 due to omitted responses or to more than 100 due to rounding.

Note 2: (+)="positive impact"
(−)="negative impact"
(0)="non-existent"